주제 5

Analysis of CoVariance

Introduction : Backgrounds

- Types of factor in Cohort study
  - a numerical outcome factor : dependent var.
  - a categorical treatment (or risk) factor
  - a numerical confounding factor : covariate
- General Linear Model(GLM)
  - can be applied to the problem of estimating treatment effects
  - represents the outcome value as a linear combination(or weighted sum) of measured variables

- Types of GLM
  - Regression model
    - the measured variables are all numerical
    - is used to model relationships between variables, estimate the form of a relationship between a response variables and a numerical inputs
  - Analysis of Variance (ANOVA)
    - the variables are all categorical
    - a technique for comparing the means of two or more populations on the basis of samples from each.
  - ANCOVA
    - represents the main application of the linear model for the purpose under the conditions with a categorical treatment factor and a numerical confounding factor

- Understandings of ANOVA
  - divide the total sample variance into within-group and between-group components
    - Within-group component=estimates of error variance (provides standard error for the inference of the group differences)
    - Between-group component=estimates of (error variance + a function of the differences among treatment means)
  - test statistic = ratio of between- to within-group variance
    - Null hypothesis : all means are equal

- Understandings of ANCOVA
  - Combining regression and ANOVA : R.A. Fisher(1932)
  - provides the powerful advantage of making possible comparisons among treatment groups differing prior to treatment
  - Assumptions :
    - a variable X that is related to the outcome Y
    - treatment groups have different means
    - assume for simplicity that X is the only variable on which the group differ
  - If we knew the relationship between Y and X, we could appropriately adjust the observed differences on Y to take account of the differences on X

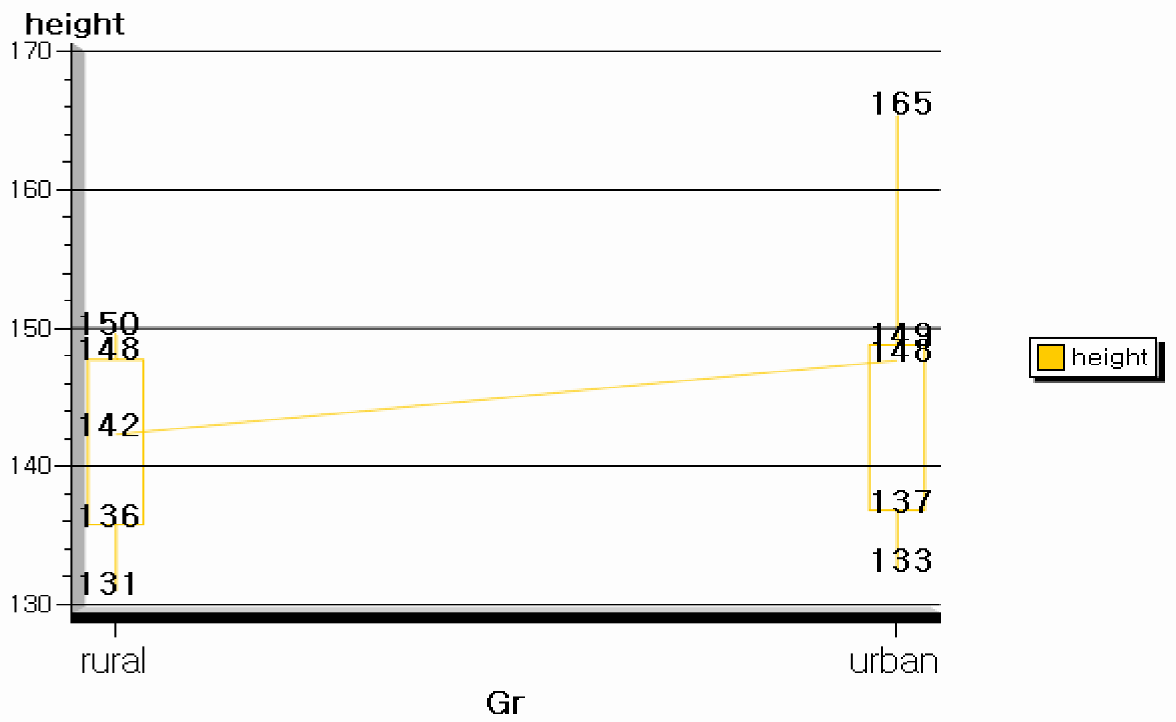
Example : Comparing the nutrition of the urban and rural children

- Greenberg(1953) : a nutrition study designed to compare growth of children in an urban with that of rural children
  - Interesting observations : heights

urban private school						rural public school					
id	age	height	id	age	height	id	age	height	id	age	height
1	109	137.6	10	129	148.3	1	121	139.0	10	138	147.7
2	113	147.8	11	130	147.5	2	121	140.9	11	138	147.7
3	115	136.8	12	133	148.8	3	128	134.9	12	140	134.6
4	116	140.7	13	134	133.2	4	129	149.5	13	140	135.8
5	119	132.7	14	135	148.7	5	131	148.7	14	140	148.5
6	120	145.4	15	137	152.0	6	132	131.0			
7	121	135.0	16	139	150.6	7	133	142.3			
8	124	133.0	17	141	165.3	8	134	139.9			
9	126	148.5	18	142	149.3	9	138	142.9			
			mean	126.8	144.5				mean	133.1	141.7

AGE : confounding factor

- **지역 아동의 평균 신장 비교**



Output : t-test and ANOVA

Statistics

Variable	Gr	N	Lower CL Mean	Mean	Upper CL Mean	Std Err	Min	Max
height	rural	14	138.14	141.67	145.21	1.6358	131	149.5
height	urban	18	140.25	144.51	148.77	2.0188	132.7	165.3
height	Diff		-8.373	-2.84	2.6935	2.7093		

T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
height	Pooled	Equal	30	-1.05	0.3030
height	Satterthwaite	Unequal	29.8	-1.09	0.2832

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
height	Folded F	17	13	1.96	0.2240

- **Output : ANOVA**

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	63.502401	63.502401	1.10	0.3030
Error	30	1734.166349	57.805545		
Corrected Total	31	1797.668750			

도시와 농촌지역의 어린이 키 차이가 없다는 귀무가설 채택

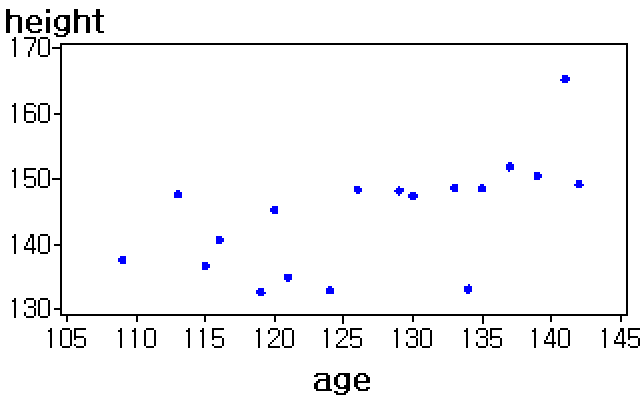
R-Square	Coeff Var	Root MSE	height Mean
0.035325	5.306807	7.602996	143.2688

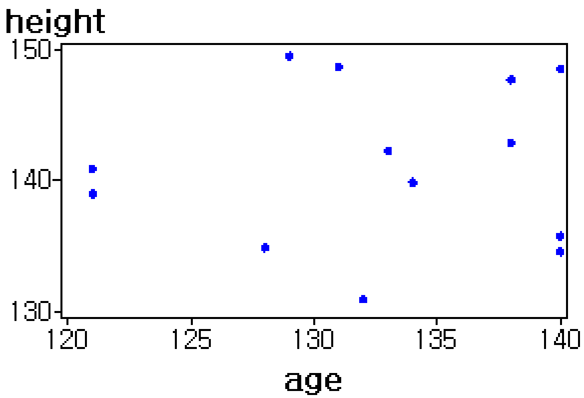
Bartlett's Test for Homogeneity of height Variance

Source	DF	Chi-Square	Pr > ChiSq
Gr	1	1.5351	0.2153

- **Explanations** :
  - the observed difference between the groups is not statistically significant
  - the mean age for the rural children is 6.3 months greater than that of the urban children
  - if the age distributions were the same, the difference in average height between the groups would be even larger than the observed 2.8
- **Alternative method : ANCOVA**
  - Need to adjust the current difference to obtain a better (or less-biased) estimate of the difference between groups that would have been observed had the mean ages in two groups been equal

- **the combination of Regression and ANOVA**
  - ANOVA : the within-group variance to reflect only random error
  - Regression : used to remove that part of the error attributable to X(covariate) and thereby to increase the precision of group comparisons





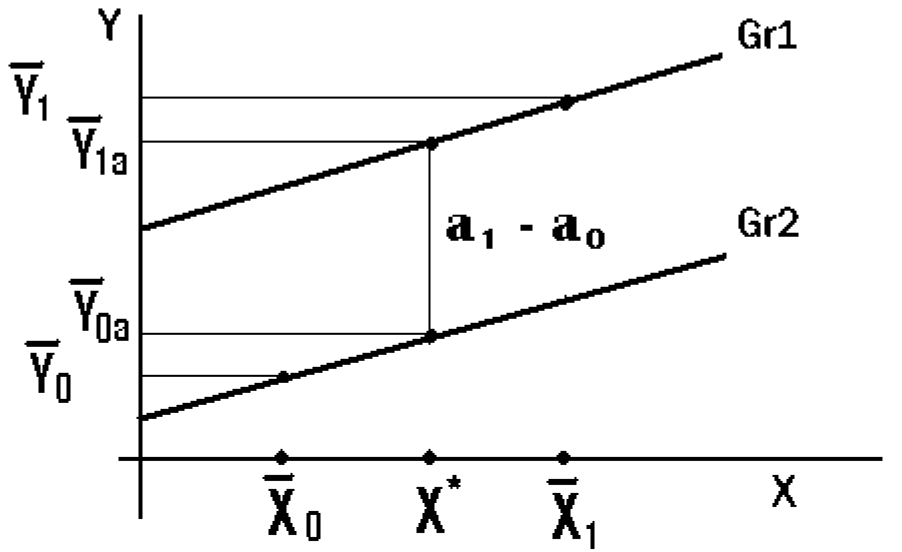
Height and age is quite linear

- **Two important benefits of ANCOVA**
  1) By estimating the form of the relationship between outcome and covariates, an appropriate adjustment can be made to remove biases resulting from group differences on the covariates
     - ✓ It is of importance primarily in randomized studies
  2) By reducing the variation within groups, the precision of estimates and tests used to compare groups can be increased
- A powerful tool for estimating treatment effects

General ANCOVA model

- Basic model
  - the use of the standard ANCOVA asserts that there is a linear relationship between the outcome Y and the covariate X with identical slopes in the two groups, but possibly different intercepts
    - Gr1(treatment) : $Y = \alpha_1 + \beta X + e$
    - Gr2(control) : $Y = \alpha_0 + \beta X + e$
      - α : the expected value of Y when X=0
      - e : random variable representing error
- Note : the direct comparison of $\overline{Y_1}$ and $\overline{Y_0}$ will be biased since $\overline{X}_1 \neq \overline{X}_0$.

- Standard ANCOVA assumption



  - GR1 : $\overline{Y_1} = \alpha_1 + \beta\ \overline{X}_1 + \bar{e}_1$
  - GR2 : $\overline{Y_0} = \alpha_0 + \beta\ \overline{X}_0 + \bar{e}_0$

$$\Rightarrow \quad E\left(\overline{Y_1} - \overline{Y_0}\right) = (\alpha_1 - \alpha_0) + \beta\left(\overline{X}_1 - \overline{X}_0\right)$$

- Note :
  - $\alpha_1 - \alpha_0$: the expected effect difference between the outcomes of the two individuals with the same value of X but in two different groups
    - To estimate $\alpha_1 - \alpha_0$, we cannot simply subtract $\overline{Y}_0$ from $\overline{Y}_1$
    - Must adjust each of these to move them to a common X(*)
      - adjusted mean of Y $\overline{Y}_{1a} = \overline{Y}_1 - \beta\left(\overline{X}_1 - X^*\right)$

        $\overline{Y}_{0a} = \overline{Y}_0 - \beta\left(\overline{X}_0 - X^*\right)$
- To estimate the difference between the means of two groups at the same value of X(*), we can simply take the difference of these two adjusted means : (unbiased estimator of $\alpha_1 - \alpha_0$)

$$\overline{Y}_{1a} - \overline{Y}_{0a} = \overline{Y}_1 - \beta\left(\overline{X}_1 - X^*\right) - \left\{\overline{Y}_0 - \beta\left(\overline{X}_0 - X^*\right)\right\}$$

$$= \left(\overline{Y}_1 - \overline{Y}_0\right) - \beta\left(\overline{X}_1 - \overline{X}_0\right)$$

- ANCOVA provides with an unbiased estimator based on the relationship between Y and X within the two groups.
  - the adjusted difference is $\bar{Y}_{1a} - \bar{Y}_{0a} = (\bar{Y}_1 - \bar{Y}_0) - \hat{\beta}(\bar{X}_1 - \bar{X}_0)$
    - where $\hat{\beta}$ is a pooled within-group regression coefficient under an assumption that the regression lines are parallel
    - Data are combined on the relationship between Y and X in both groups
- Create the overall regression line fitted to the entire sample
  - if incorrect, it does not yield an unbiased estimate of β or of the effect $\alpha_1 - \alpha_0$ under the model

Assumptions underlying

- Equality of regression slopes
- Linearity
- Covariate measured without error
- No unmeasured confounding variables
- Errors independent of each other
- Equality of error variance
- Normality of errors

Combined ANCOVA model

- Combined ANCOVA model with K covariates and J treatments/groups
  - the general ANCOVA model (in SAS)

$$y_{ij} = \mu + \alpha_i + \beta_1\left(x_{1ij} - \bar{x}_{1\bullet\bullet}\right) + \cdots + \beta_K\left(x_{Kij} - \bar{x}_{K\bullet\bullet}\right) + \varepsilon_{ij} \quad \left(i = 1, \cdots, I\right)$$

$$= \beta_0 + \alpha_i + \beta_1 x_{1ij} + \cdots + \beta_K x_{Kij} + \varepsilon_{ij}$$

$$= \beta_{0i} + \beta_1 x_{1ij} + \cdots + \beta_K x_{Kij} + \varepsilon_{ij}$$

$$\textit{where} \quad \beta_0 = \mu - \beta_1\bar{x}_{1\bullet\bullet} - \cdots - \beta_K\bar{x}_{K\bullet\bullet} \text{ and } \quad \beta_{0i} = \beta_0 + \alpha_i$$

$$\alpha_i \ : \text{the effect of } i\textit{th} \text{ treatmen ts/groups}$$

- Two tests in ANCOVA
  - Test of the effect of covariates

$$H_0 : \beta_i = 0$$

    - If null hypothesis is accepted, we use ANOVA instead of ANCOVA
  - Test of the effect of treatments/groups

$$H_0 : \text{ all } \alpha_i \text{ are equal or } \alpha_1 = \cdots = \alpha_I$$

- Anova table of ANCOVA

Source	df	SS	MS	F
Covariate1	1	SSX1	MSX1=SSX1/1	F1=MSX1/MSE
……	……	……	……	……
CovariateK	1	SSXK	MSXK=SSXK/1	FK=MSXK/MSE
Treatment	I-1	SStr	MStr=SSTR/(I-1)	Ftr=MStr/MSE
Error	N-(K-1)-(I-1)	SSE	MSE=SSE/dfE	
Total	N-1	SST		

  - tests of covariates $$F_i = \frac{SSX_i / 1}{SSE / df_E} \sim F(1, df_E)$$
  - test of treatment $$F_{Tr} = \frac{SStr / (I-1)}{SSE / df_E} \sim F(I-1, df_E)$$

- 제곱합의 표현 : 범주형 및 공변량이 각각 하나인 경우

$$SST = \sum_{i=1}^{I} \sum_{j=1}^{n_i} \left(y_{ij} - \bar{y} \right)^2$$

$$SStr = \frac{S_{xx} S_{yy} - \left(S_{xy} \right)^2}{S_{xx}} - SSE \quad , \quad SSX = \sum_{i=1}^{I} S_{yy(i)} - SSE$$

$$SSE = \frac{\sum_{i=1}^{I} S_{xx(i)} \sum_{i=1}^{I} S_{yy(i)} - \left(\sum_{i=1}^{I} S_{xy(i)} \right)^2}{\sum_{i=1}^{I} S_{xx(i)}}$$

  - where

$$S_{xx(i)} = \sum_{j=1}^{n_i} \left(x_{ij} - \bar{x}_i \right)^2 \quad , \quad S_{yy(i)} = \sum_{j=1}^{n_i} \left(y_{ij} - \bar{y}_i \right)^2$$

$$S_{yy(i)} = \sum_{j=1}^{n_i} \left(x_{ij} - \bar{x}_i \right) \left(y_{ij} - \bar{y}_i \right)$$

$$S_{xx} = \sum_{i=1}^{I} \sum_{j=1}^{n_i} \left(x_{ij} - \bar{x} \right)^2 \quad , \quad S_{yy} = \sum_{i=1}^{I} \sum_{j=1}^{n_i} \left(y_{ij} - \bar{y} \right)^2 \quad ,$$

$$S_{xy} = \sum_{i=1}^{I} \sum_{j=1}^{n_i} \left(x_{ij} - \bar{x} \right) \left(y_{ij} - \bar{y} \right)$$

Example : (continued)

- **Using ANCOVA,**
  - the slope between height and age (LSE)

$$\hat{\beta} = 0.4118$$

  - the mean age of groups

$$\bar{X}_1 = 126.8 \quad \text{and} \quad \bar{X}_0 = 133.1$$

  - the difference of both groups before adjustment

$$\bar{Y}_1 - \bar{Y}_0 = 144.5 - 141.7 = 2.8$$

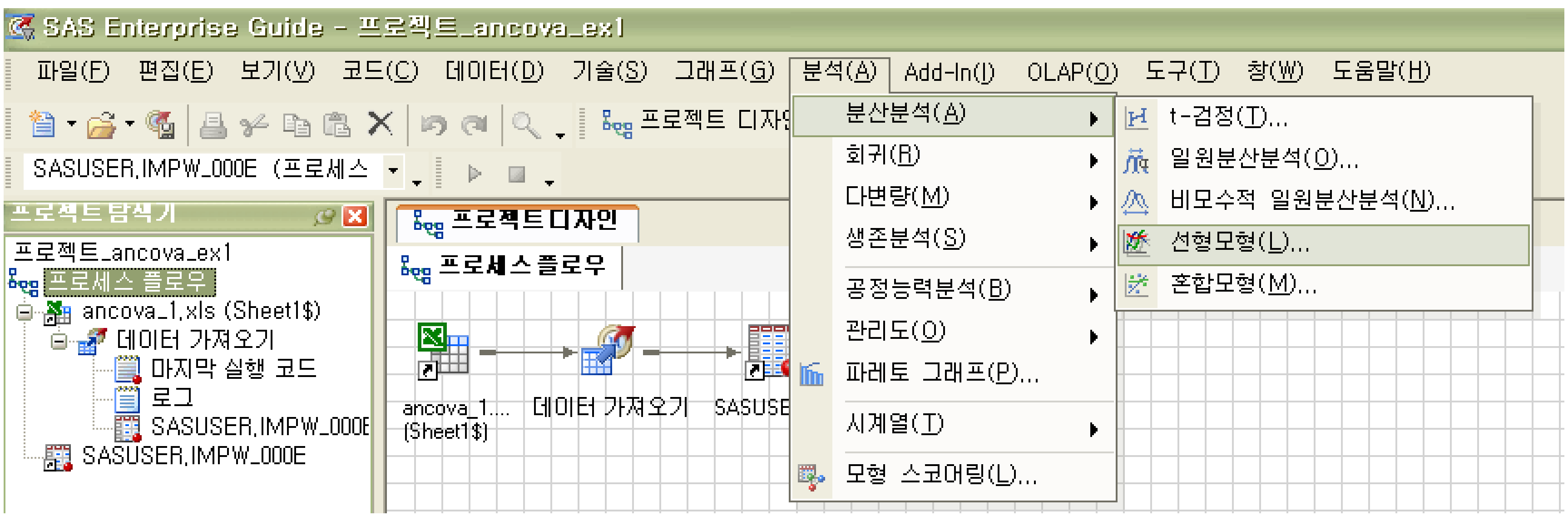
  - the adjusted difference

$$\bar{Y}_{1a} - \bar{Y}_{0a} = \left(\bar{Y}_1 - \bar{Y}_0\right) - \hat{\beta}\left(\bar{X}_1 - \bar{X}_0\right)$$

$$= 2.8 - 0.4118 \times \left(126.8 - 133.1\right) = 5.4$$

- **We may ask whether the adjusted difference is statistically significant , $H_0 : \alpha_1 = \alpha_0$**

- **SAS procedures for ANCOVA**
  - proc GLM : "분산분석"의 선형모형
  - proc Mixed : "분산분석"의 혼합모형
  - proc Genmod : "회귀"의 일반화선형모형




IMPW_000E에 대한 선형모형
작업 역할
모형
모형 옵션
고급 옵션
Post Hoc 검정
최소제곱
산술
도표
평균
예측값
잔차
영향력
예측값
제목
작업 역할
할당할 변수(A):
이름
Gr
id
age
height
작업 역할(E):
종속변수 (제한: 1개)
height
양적 변수
age
분류변수
Gr
그룹 분석변수
빈도변수 (제한: 1개)
상대 가중값 변수 (제한: 1개)
IMPW_000E에 대한 선형모형
작업 역할
모형
모형 옵션
고급 옵션
Post Hoc 검정
최소제곱
산술
도표
평균
예측값
잔차
영향력
예측값
제목
모형
분류변수 및 양적변수(V):
age
Gr
주효과(M)
교차(O)
지분효과(N)
계승(F)
차수(D):
다항(P)
차수(G):
2
효과(E):
age
Gr

- **SAS output : GLM / Mixed / GENMOD**

Source	DF	**Sum of Squares**	Mean Square	F Value	Pr > F
Model	2	458.208058	229.104029	4.96	0.0140
Error	29	1339.460692	46.188300		
Corrected Total	31	1797.668750			

Source	DF	**Type I SS**	Mean Square	F Value	Pr > F
age	1	254.6436624	254.6436624	5.51	0.0259
Gr	1	203.5643951	203.5643951	4.41	0.0446

Source	DF	**Type III SS**	Mean Square	F Value	Pr > F
age	1	394.7056568	394.7056568	8.55	0.0067
Gr	1	203.5643951	203.5643951	4.41	0.0446

- **Age and Gr : both are statistically significant**

If not reject,
Use ANOVA instead of ANCOVA

- **Estimation of coefficients**

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	92.27968294	B	17.93905749	5.14	<.0001
age	0.41181152		0.14087301	2.92	0.0067
Gr rural	-5.40860204	B	2.57632179	-2.10	0.0446
Gr urban	0.00000000	B	.	.	.

  - **adjusted mean is the coefficient of Gr:rural**

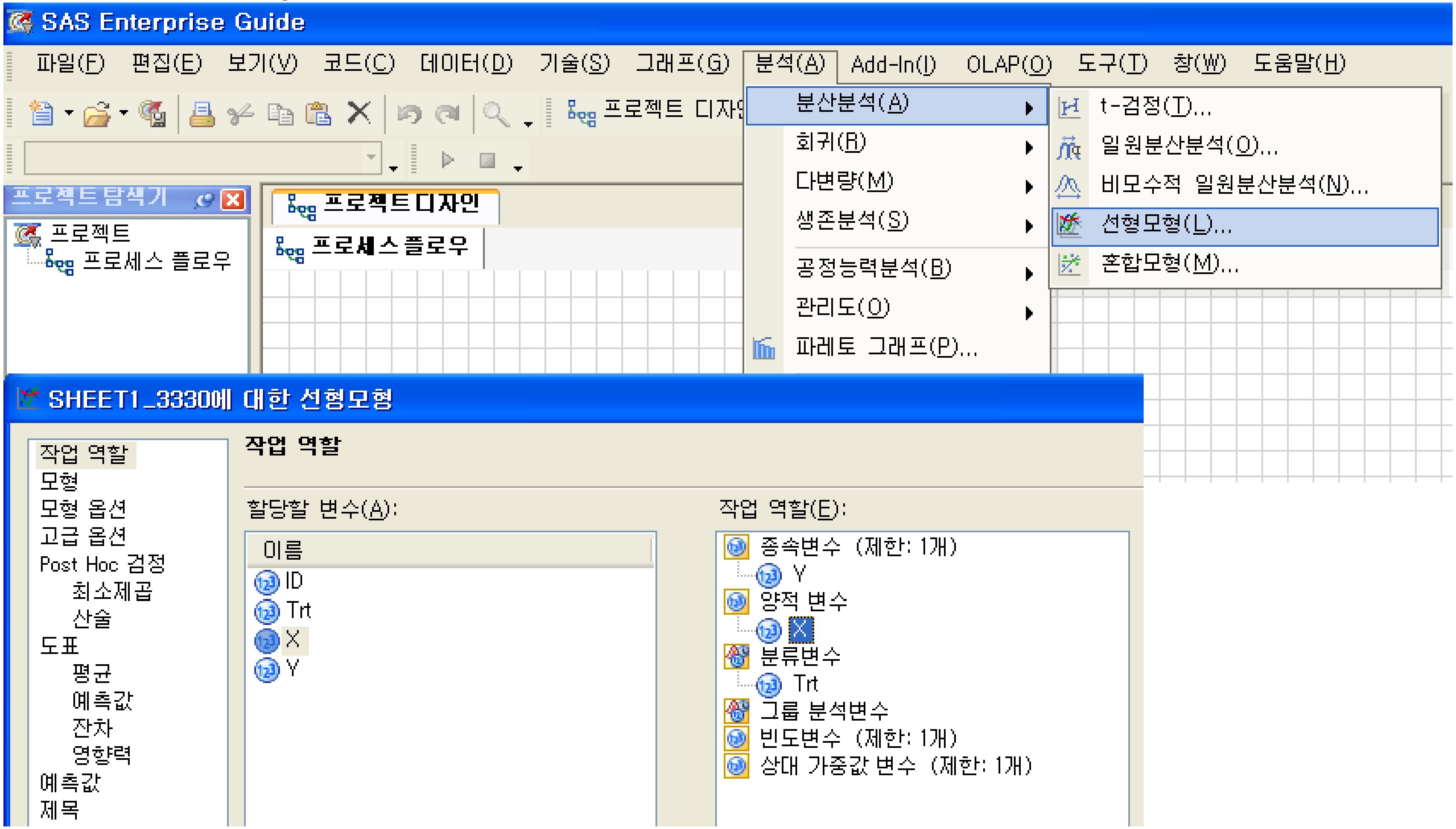
- **age : statistically significant**
- **Gr : statistically significant**
  - **the coef. is the estimate of the difference of effect for the treatment group**

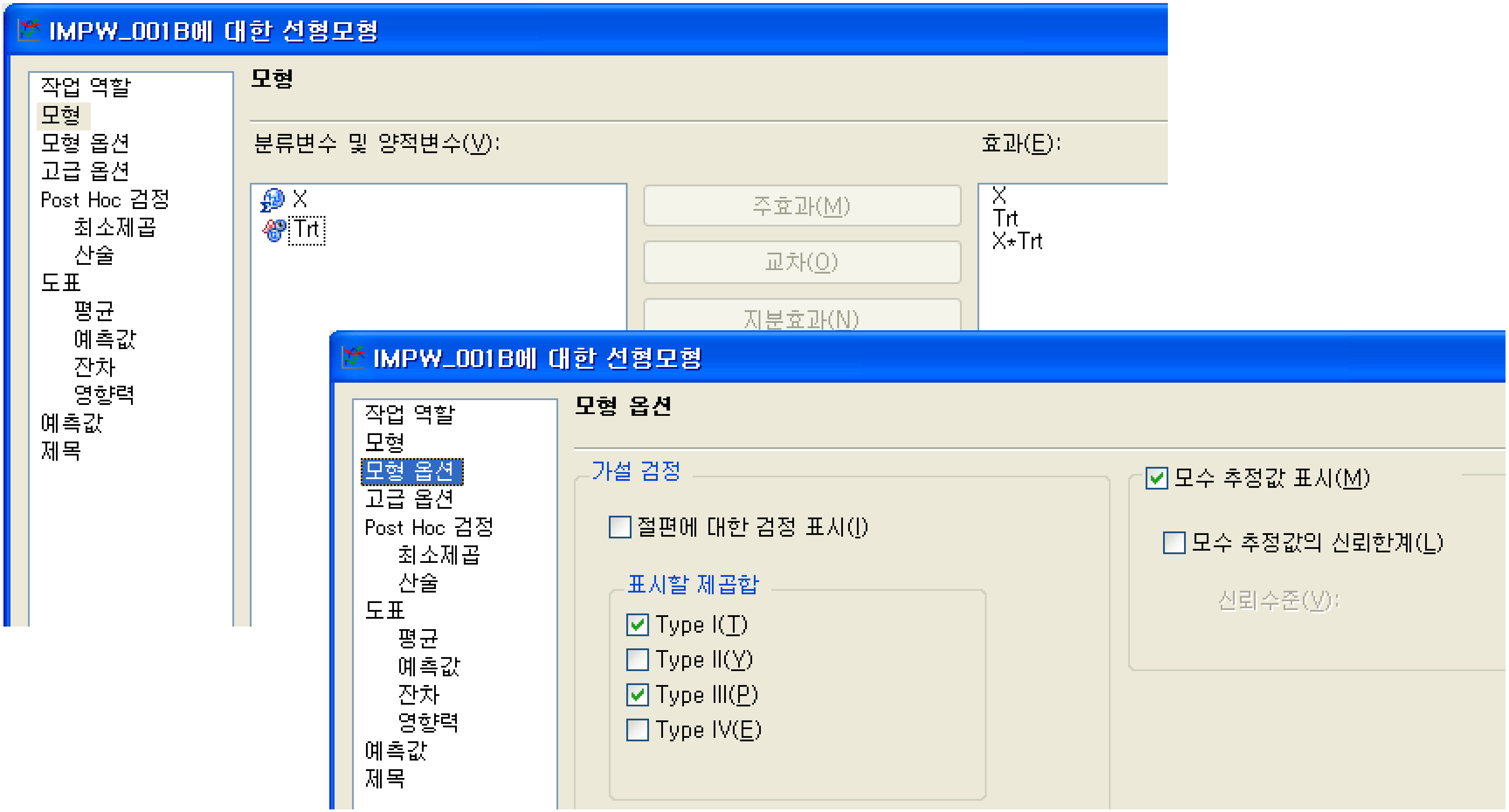
Example 2 : Hypoglycemia

- **Study of hypoglycemia : Comparison with 5 treatments** (textbook, p.p. 111)
  - participants are 20, 5 random allocations are consisted
  - **interesting variable : value of blood sugar**
  - covariate variable : value of the beginning blood sugar

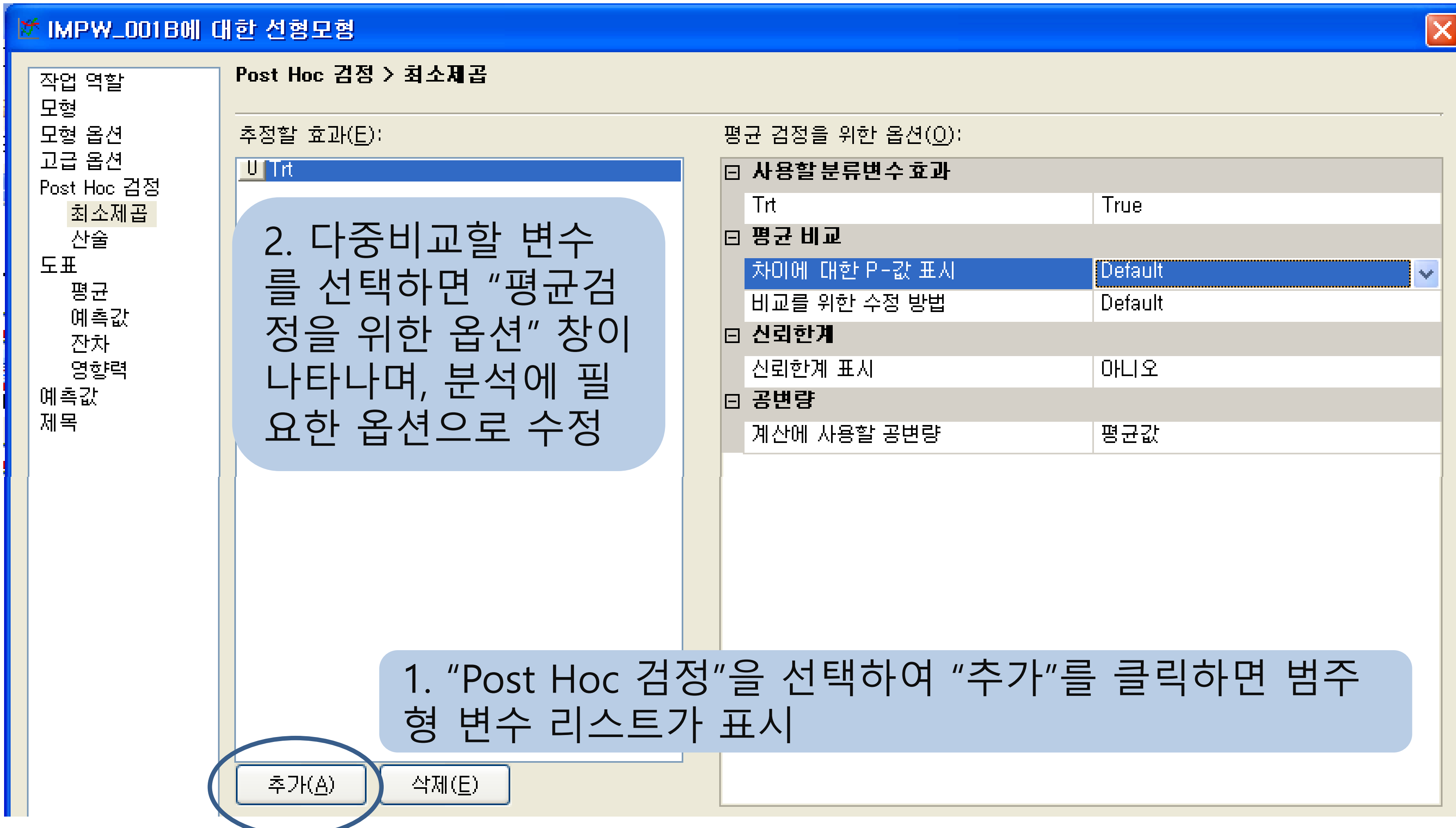
ID	Trt 1		Trt 2		Trt 3		Trt 4		Trt 5	
	X	Y	X	Y	X	Y	X	Y	X	Y
1	27.2	32.6	28.6	33.8	28.6	35.2	29.3	35.0	20.4	24.6
2	22.0	36.6	26.8	31.7	22.4	29.1	21.8	27.0	19.6	23.4
3	33.0	37.7	26.5	30.7	23.2	28.9	30.3	36.4	25.1	30.3
4	26.8	31.0	26.8	30.4	24.4	30.2	24.3	30.5	18.1	21.8

- SAS EG
  - SAS : proc GLM 이용




IMPW_001B에 대한 선형모형
작업 역할
모형
모형 옵션
고급 옵션
Post Hoc 검정
최소제곱
산술
도표
평균
예측값
잔차
영향력
예측값
제목
모형
분류변수 및 양적변수(V):
X
Trt
주효과(M)
교차(O)
지분효과(N)
효과(E):
X
Trt
X*Trt
IMPW_001B에 대한 선형모형
작업 역할
모형
모형 옵션
고급 옵션
Post Hoc 검정
최소제곱
산술
도표
평균
예측값
잔차
영향력
예측값
제목
모형 옵션
가설 검정
절편에 대한 검정 표시(I)
표시할 제곱합
Type I(T)
Type II(Y)
Type III(P)
Type IV(E)
모수 추정값 표시(M)
모수 추정값의 신뢰한계(L)
신뢰수준(V):

• 다중비교/사후분석 : 수준간 평균 차이 분석



2. 다중비교할 변수를 선택하면 "평균검정을 위한 옵션" 창이 나타나며, 분석에 필요한 옵션으로 수정

1. "Post Hoc 검정"을 선택하여 "추가"를 클릭하면 범주형 변수 리스트가 표시

- Output of the saturated model

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	327.4111358	36.3790151	11.64	0.0003
Error	10	31.2583642	3.1258364		
Corrected Total	19	358.6695000			

Source	DF	**Type I SS**	Mean Square	F Value	Pr > F
X	1	256.7473944	256.7473944	82.14	<.0001
Trt	4	34.1881007	8.5470252	2.73	0.0898
X*Trt	4	36.4756407	9.1189102	2.92	0.0773

Source	DF	**Type III SS**	Mean Square	F Value	Pr > F
X	1	51.98115809	51.98115809	16.63	0.0022
Trt	4	44.45642772	11.11410693	3.56	0.0472
X*Trt	4	36.47564068	9.11891017	2.92	0.0773

[공변량 효과의 동일성 검정] 포화모형으로 분석한 결과, 공변량과 처리간 교호작용이 존재하지 않는다는 귀무가설이 유의수준 5%에서 채택되므로 교호작용을 제외한 공분산분석으로 분석해야 함을 보여줌

- Output of ANCOVA model

1. 모형 적합성 : 분산분석표
   - 결론 : 유의수준 5%에서 공분산 모형은 적합하다는 결과를 보여줌

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	290.9354951	58.1870990	12.03	0.0001
Error	14	67.7340049	4.8381432		
Corrected Total	19	358.6695000			

- 혈당량 수치에 대한 공분산 모형의 설명력은 81% 정도임

R-Square	Coeff Var	Root MSE	Y Mean
0.811152	7.131068	2.199578	30.84500

- Types of Sum of squares

- 제 1종 제곱합 (순차 제곱합) : 모형에 추가되는 순서에 의해 계산되는 순차제곱합

$$SS \quad of \quad \text{model} \quad = SS\,(trt\,,X\,) = SS\,(\,X\,) + SS\,(trt\,\mid X\,)$$

- 모형식에 추가되는 순서에 따라 제곱합이 다르게 표현됨(교재와 비교 요망)

Source	DF	**Type I SS**	Mean Square	F Value	Pr > F
X	1	256.7473944	256.7473944	53.07	<.0001
Trt	4	34.1881007	8.5470252	1.77	0.1917

- 제 3종 제곱합 (고유효과 제곱합) : 개별 요인이 종속변수에 순수하게 미치는 고유 기여분에 대한 제곱합 (모형 진입순서에 관계없이 항상 동일)

$$SS\,3 = SS\,(\,X\,\mid trt\,) + SS\,(trt\,\mid X\,)$$

Source	DF	**Type III SS**	Mean Square	F Value	Pr > F
X	1	92.52849508	92.52849508	19.12	0.0006
Trt	4	34.18810071	8.54702518	1.77	0.1917

- 공분산 분석에서는 공변량 효과를 제어한 범주형 효과의 차이를 분석하므로 X(공변량)의 영향력을 보정한 (trt|X)의 분석이 관심 대상
  -- 제 3종 제곱합을 이용하여 효과에 대한 분석 수행

- **cf. : Types of Sum of Squares**
  - **Ex. : Two-way ANOVA** $R(\alpha, \beta, \alpha\beta \mid \mu)$ **를 가정**
    - **분할 순서 :** $\alpha \rightarrow \beta \rightarrow \alpha\beta$

$$R(\alpha, \beta, \alpha\beta \mid \mu) = R(\alpha \mid \mu) + R(\beta \mid \mu, \alpha) + R(\alpha\beta \mid \mu, \alpha, \beta)$$

효과	Type I SS	Type II SS	Type III SS
α	$R(\alpha \mid \mu)$	$R(\alpha \mid \mu, \beta)$	$R(\alpha \mid \mu, \beta, \alpha\beta)$
β	$R(\beta \mid \mu, \alpha)$	$R(\beta \mid \mu, \alpha)$	$R(\beta \mid \mu, \alpha, \alpha\beta)$
$\alpha\beta$	$R(\alpha\beta \mid \mu, \alpha, \beta)$	$R(\alpha\beta \mid \mu, \alpha, \beta)$	$R(\alpha\beta \mid \mu, \alpha, \beta)$

- **여기서 type IV SS 는 모든** $n_{ij} > 0$ **인 경우 Type III SS 와 동일**
- **Model SS의 표현**

$$\text{Model SS} = \sum \text{Type I SS} = R(\alpha \mid \mu) + R(\beta \mid \mu, \alpha) + R(\alpha\beta \mid \mu, \alpha, \beta)$$

Estimates of parameters and equation

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	9.353378171	B	3.74853167	2.50	0.0257
X	0.753443357		0.17228688	4.37	0.0006
Trt 1	4.590290346	B	1.91153056	2.40	0.0308
Trt 2	1.821798598	B	1.90404778	0.96	0.3549
Trt 3	2.924243075	B	1.69087092	1.73	0.1057
Trt 4	2.961881116	B	1.83255369	1.62	0.1283
Trt 5	0.000000000	B	.	.	.

- 모형식
  - Meaning that estimates of Trt5=0 : control group
  - Estimate of the covariate X is equal
  - Estimate of the trt's are the differences between trt5 and trt(i)
  - Equations of Trt's

$$Trt\ 1:\ \hat{y}_{1j} = \hat{\beta}_0 + \hat{\alpha}_1 + \hat{\beta}\, x_{1j} = 9.353 + 4.59 + 0.7534\ x_{1j}$$

$$\vdots$$

$$Trt\ 5:\ \hat{y}_{5j} = \hat{\beta}_0 + \hat{\alpha}_5 + \hat{\beta}\, x_{5j} = 9.353 + 0 + 0.7534\ x_{1j}$$

- Least square mean :
  - 처리수준의 비교를 위해 공변량(X) 전체 평균을 대입하여 계산된 반응변수(결과)의 평균
    - 공변량 효과를 보정한 반응변수의 평균
    - 공변량 분석의 주 분석 관심 대상
  - LSMEAN의 계산

$$\bar{y}_{1adj} = \hat{\beta}_0 + \hat{\alpha}_1 + \hat{\beta}\,\bar{x} = 9.353 + 4.59 + .7534\,(25.26) = 32.975$$

Trt	Y LSMEAN	LSMEAN Number
1	32.9756477	1
2	30.2071560	2
3	31.3096004	3
4	31.3472385	4
5	28.3853574	5

- Multiple comparisons :
  - 공변량 효과를 보정한 처리별 반응변수(결과) 평균의 차이에 대한 다중비교를 의미
  - SAS EG에서는 처리 수준간 차이에 대한 유의확률(p-값)을 제시해 줌

귀무가설

Least Squares Means for effect Trt
Pr > |t| for H0: LSMean(i)=LSMean(j)
Dependent Variable: Y

i/j	1	2	3	4	5
1		0.0968	0.3208	0.3148	0.0308
2	0.0968		0.5060	0.4771	0.3549
3	0.3208	0.5060		0.9814	0.1057
4	0.3148	0.4771	0.9814		0.1283
5	0.0308	0.3549	0.1057	0.1283	

- 처리간 평균의 차이가 없다는 가설에 대해 유의수준 5%보다 작은 경우 귀무가설은 기각하며, 두 수준간 평균에 차이가 있음을 의미

Example 3 : Medicament

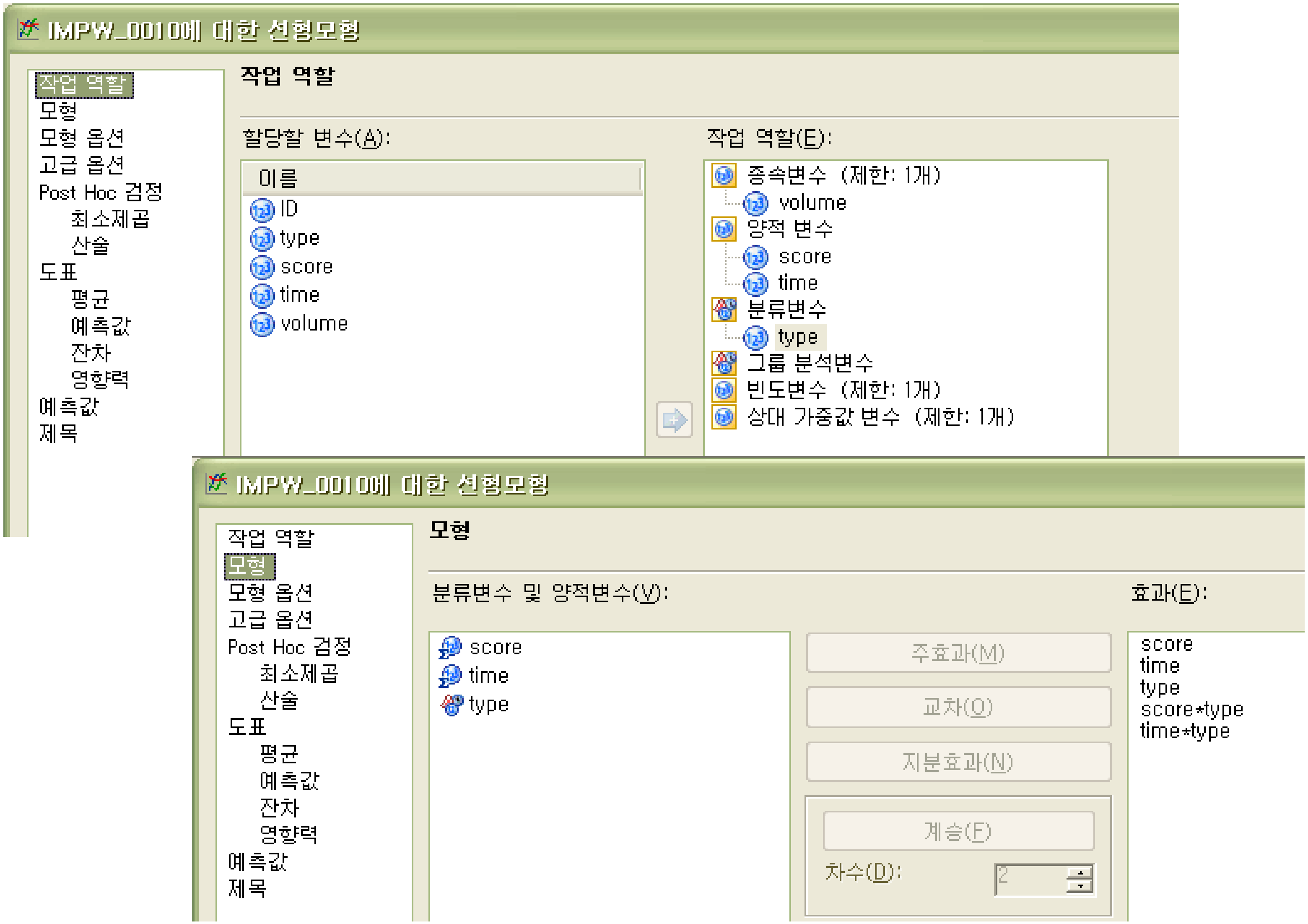
- **Elimination/약물에 대한 체내배출 연구 :**
  - covariates : anti-metabolic score(X1), time(X2)
  - categorical variable : type of drug
  - interesting variable : volume of medicament

id	type 1			type 2			type 3			type 4			type 5			type 6		
	sc	t	volume	sc	t	volume	sc	t	volume	sc	t	volume	sc	t	volume	sc	t	volume
1	37	61	11.3208	37	37	12.9151	45	53	18.8947	41	41	14.6739	57	41	8.6493	49	33	9.5238
2	49	49	7.6923	53	53	0.0017	53	45	8.0477	53	53	6.7358	61	37	6.1441	49	65	21.7939
3	53	45	4.2553	57	57	0.0017	49	49	11.0196	53	53	6.2762	53	45	13.2316	53	53	5.0676
4	57	57	5.6235	49	49	14.9893	53	53	13.7233	53	45	6.0669	53	53	8.1602	61	37	1.4423
5	45	45	6.9971	53	45	5.2308	57	57	8.2560	49	49	14.5000	53	53	20.7627	53	45	3.6115
6	53	53	11.3475	53	45	9.4650	53	53	22.6103	61	37	0.0020	49	65	20.5997	37	37	28.1828

  - 모형식

$$
\begin{aligned}
y_{ij} &= \mu + \alpha_i + \beta_1\left(x_{1ij} - \bar{x}_1\right) + \beta_2\left(x_{2ij} - \bar{x}_2\right) + \varepsilon_{ij} \\
&= \beta_0 + \alpha_i + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \varepsilon_{ij}
\end{aligned}
$$

$$
\text{where} \quad \beta_0 = \mu - \beta_1 \bar{x}_1 - \beta_2 \bar{x}_2 \;,\; \varepsilon_{ij} \sim N(0, \sigma^2)
$$


IMPW_0010에 대한 선형모형
작업 역할
모형
모형 옵션
고급 옵션
Post Hoc 검정
최소제곱
산술
도표
평균
예측값
잔차
영향력
예측값
제목
작업 역할
할당할 변수(A):
이름
ID
type
score
time
volume
작업 역할(E):
종속변수 (제한: 1개)
volume
양적 변수
score
time
분류변수
type
그룹 분석변수
빈도변수 (제한: 1개)
상대 가중값 변수 (제한: 1개)
IMPW_0010에 대한 선형모형
작업 역할
모형
모형 옵션
고급 옵션
Post Hoc 검정
최소제곱
산술
도표
평균
예측값
잔차
영향력
예측값
제목
모형
분류변수 및 양적변수(V):
score
time
type
주효과(M)
교차(O)
지분효과(N)
계승(F)
차수(D):
2
효과(E):
score
time
type
score*type
time*type

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	17	1204.578049	70.857532	2.97	0.0136
Error	18	429.097741	23.838763		
Corrected Total	35	1633.675790			

Source	DF	**Type III SS**	Mean Square	F Value	Pr > F
score	1	135.1152440	135.1152440	5.67	0.0285
time	1	19.9537201	19.9537201	0.84	0.3723
type	5	99.2244578	19.8448916	0.83	0.5434
score*type	5	172.6899491	34.5379898	1.45	0.2550
time*type	5	42.3296596	8.4659319	0.36	0.8722

- [공분산 모형 분석을 위한 검정 : 약의 형태에 관계없이 공변량의 영향은 동일한가에 대한 검정] 공변량과 약의 형태에 대한 교호작용에 대한 가설(교호작용의 영향은 유의하지 않다)의 검정을 통해 귀무가설이 채택되면 교호작용이 유의하지 않으므로 모든 처리에 대해 기울기가 동일함을 의미하고, 공분산분석으로 분석 가능함을 의미

- **공분산 모형의 유의성/적합성 :**

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	1001.534875	143.076411	6.34	0.0002
Error	28	632.140915	22.576461		
Corrected Total	35	1633.675790			

- **개별 효과에 대한 유의성/영향력 :**
  - **공변량 효과 보정 전 : SS1**

Source	DF	**Type I SS**	Mean Square	F Value	Pr > F
score	1	516.5921408	516.5921408	22.88	<.0001
time	1	62.7261445	62.7261445	2.78	0.1067
type	5	422.2165894	84.4433179	3.74	0.0102

  - **공변량 효과 보정 후 : SS3**

Source	DF	**Type III SS**	Mean Square	F Value	Pr > F
score	1	691.2448912	691.2448912	30.62	<.0001
time	1	54.4043164	54.4043164	2.41	0.1318
type	5	422.2165894	84.4433179	3.74	0.0102

- **Estimates of parameters**

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	42.48130557	B	8.68108395	4.89	<.0001
score	-0.76073579		0.13748199	-5.53	<.0001
time	0.16472695		0.10611479	1.55	0.1318
type 1	-5.84339403	B	2.83830488	-2.06	0.0489
type 2	-4.94232186	B	2.75781810	-1.79	0.0839
type 3	2.07108475	B	2.83956405	0.73	0.4718
type 4	-2.76650487	B	2.75314779	-1.00	0.3236
type 5	3.70498538	B	2.83099062	1.31	0.2013
type 6	0.00000000	B	.	.	.

  - **Type 의 추정치는 기준범주 6과의 차이를 의미**
  - **약의 형태별로 공분산 모형식 표현 : 공변량 기울기는 모두 동일**
- **모형식**

$$Type\ 1:\ \hat{y}_{1j} = \hat{\beta}_0 + \hat{\alpha}_1 + \hat{\beta}\ x_{11j} + \hat{\beta}\ x_{12j} = 42.481 - 5.843 - 0.761\ score_{11j} + 0.165\ time_{12j}$$

- **Adjusted mean(LS mean)**

type	volume LSMEAN	LSMEAN Number
1	5.6697422	1
2	6.5708144	2
3	13.5842210	3
4	8.7466313	4
5	15.2181216	5
6	11.5131362	6

- 약의 형태에 대한 보정된 반응변수의 평균
  - 두 공변량 효과를 제어하고 계산
  - 공변량 전체 평균 이용

- **Multiple comparisons**

All pairwise difference

Least Squares Means for effect type
Pr > |t| for H0: LSMean(i)=LSMean(j)
Dependent Variable: volume

i/j	1	2	3	4	5	6
1		0.9995	0.0767	0.8814	0.0256	0.3367
2	0.9995		0.1522	0.9669	0.0467	0.4866
3	0.0767	0.1522		0.5263	0.9911	0.9765
4	0.8814	0.9669	0.5263		0.2176	0.9124
5	0.0256	0.0467	0.9911	0.2176		0.7778
6	0.3367	0.4866	0.9765	0.9124	0.7778	

Least Squares Means for effect type
Pr > |t| for H0: LSMean(i)=LSMean(j)
Dependent Variable: volume

i/j	1	2	3	4	5	6
1		0.7484	0.0079	0.2852	0.0023	0.0489
2	0.7484		0.0177	0.4360	0.0045	0.0839
3	0.0079	0.0177		0.0952	0.5616	0.4718
4	0.2852	0.4360	0.0952		0.0275	0.3236
5	0.0023	0.0045	0.5616	0.0275		0.2013
6	0.0489	0.0839	0.4718	0.3236	0.2013	

Default/Fisher's LSD

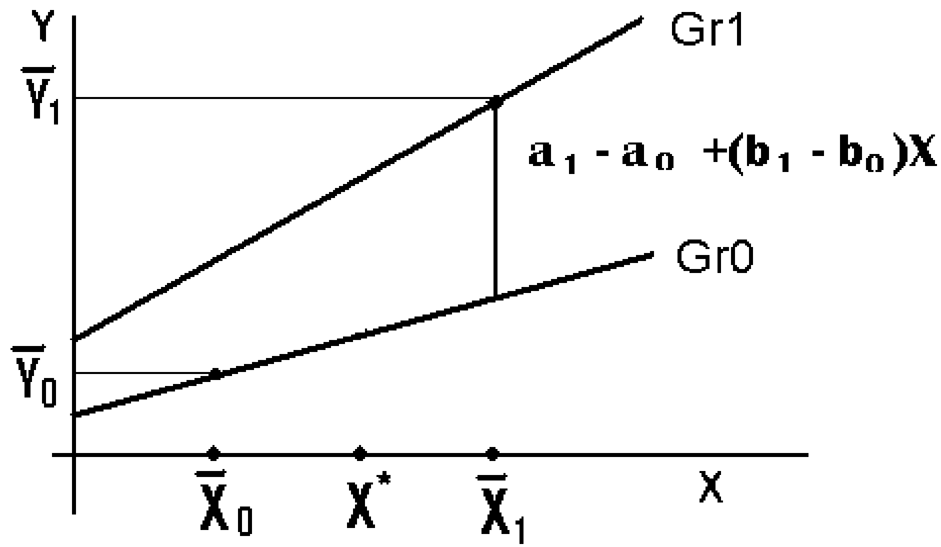
Dealing with departures from standard Assumptions

- Departures from these assumptions may result in **biased effect** estimates and/or a **loss of precision** in statistical tests and estimates

- Sources of the possibility of **bias**
  - Linearity of the relationship between Y and X
  - **Same slopes for regression lines** in two groups
  - Absence of measurement error in the covariate
  - Absence of other unmeasured covariate

Nonparallel Regression : - the different slopes

- Nonparallel linear Regression
  - **Interaction between the treatment effect and the covariate**



- the basic model $Y = \alpha_i + \beta_i X + e \quad (\text{where} \quad i = 0,1)$
- the difference between the expected outcomes with the same X : the treatment effect is a linear function of X

$$\alpha_1 - \alpha_0 + (\beta_1 - \beta_0) X$$

- Estimation of β_0 and β_1 separately from the two treatment groups

$$\overline{Y}_{1a} = \overline{Y}_1 - \hat{\beta}_1 \left(\overline{X}_1 - X\right)$$

$$\overline{Y}_{0a} = \overline{Y}_0 - \hat{\beta}_0 \left(\overline{X}_0 - X\right)$$

  - an unbiased estimate of the treatment effect for any X

$$\begin{aligned}
\overline{Y}_{1a} - \overline{Y}_{0a} &= \overline{Y}_1 - \hat{\beta}_1 \left(\overline{X}_1 - X\right) - \left\{\overline{Y}_0 - \hat{\beta}_0 \left(\overline{X}_0 - X\right)\right\} \\
&= \left(\overline{Y}_1 - \overline{Y}_0\right) - \hat{\beta}_1 \left(\overline{X}_1 - X\right) + \hat{\beta}_0 \left(\overline{X}_0 - X\right) \cdots \cdots (1) \\
&= \left(\overline{Y}_1 - \overline{Y}_0\right) - \hat{\beta}_0 \left(\overline{X}_1 - \overline{X}_0\right) \quad \text{if} \quad X = \overline{X}_1 \cdots \cdots (2)
\end{aligned}$$

  - cf. the standard ANCOVA model : replace $\hat{\beta}$

$$\overline{Y}_{1a} - \overline{Y}_{0a} = \left(\overline{Y}_1 - \overline{Y}_0\right) - \hat{\beta} \left(\overline{X}_1 - \overline{X}_0\right)$$

- **Note** :
  - Eq. (1) is more general than the usual ANCOVA model that represents the special case when $\beta_1 = \beta_0$
  - Unless $\beta_1 = \beta_0$, we estimate separate coefficients using smaller samples. This may lead to a **slight decrease in precision**.
  - We have considered for comparing treatments when regression lines are nonparallel involve specifying a particular covariate value and estimating the effect conditional on this value

- **Testing for homogeneity** of the effect of covariate between groups

$$H_0 : \text{all } \beta_{ij} \text{ of covariate } X_j \text{ are equal}$$

  - If hypothesis is rejected, we can't use ANCOVA model
  - other expression (in SAS)

$$\text{simple case} : Y = \beta_0 + \alpha_i + \beta_i X + e$$

$$= \beta_0 + \alpha_i + \left(\beta_1 + \delta_i\right) X + e$$

  - Null hypothesis : test of the relationship(interaction) between the covariate and the treatment

$$H_0 : \delta_i = 0$$

Example 4 : capability of carrying

- Data of the capability of carrying oxygen with smoker
  - interesting variable : capability of carrying oxygen
  - treatments : 3 types
  - covariate : a number of cigarette consumption per 1 day

ID	trt 1		trt 2		trt 3	
	cigar	capab	cigar	capab	cigar	capab
1	40	165	85	11	65	89
2	54	85	83	6	25	64
3	85	9	65	51	34	87
4	95	43	98	18	20	45
5	81	94	47	189	30	56
6	26	226	74	90	29	87
7	90	7	75	10	100	59
8	95	9	97	12	85	39
9	83	12	79	35	24	87
10	83	145	91	27	26	67

- **SAS output**
  - 모형의 적합성

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	63529.20505	12705.84101	10.46	<.0001
Error	24	29146.26161	1214.42757		
Corrected Total	29	92675.46667			

  - 개별 요인의 주효과 및 교호작용효과에 대한 유의성 검정

Source	DF	Type I SS	Mean Square	F Value	Pr > F
cigar	1	26520.03716	26520.03716	21.84	<.0001
trt	2	11872.60823	5936.30412	4.89	0.0166
cigar*trt	2	25136.55966	12568.27983	10.35	0.0006

Source	DF	Type III SS	Mean Square	F Value	Pr > F
cigar	1	43339.41912	43339.41912	35.69	<.0001
trt	2	34647.77546	17323.88773	14.27	<.0001
cigar*trt	2	25136.55966	12568.27983	10.35	0.0006

  - 공변량(1일 흡연량)과 치료법과의 교호작용에 대한 p-값이 0.0006이므로 치료법마다 대상의 흡연소비량의 효과가 다름
    - 치료법마다 흡연소비량의 영향이 다르므로 각 치료법마다 흡연소비량의 효과 추정이 필요함을 의미

- Estimates of parameters

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	75.6124254	B	20.8472	3.63	0.0013
cigar	-0.1737997	B	0.40403	-0.43	0.6709
trt 1	198.4239034	B	42.1708	4.71	<.0001
trt 2	208.912056	B	64.2887	3.25	0.0034
trt 3	0	B	.	.	.
cigar*trt 1	-2.4838005	B	0.62559	-3.97	0.0006
cigar*trt 2	-2.8441409	B	0.85476	-3.33	0.0028
cigar*trt 3	0	B	.	.	.

기준 범주

– 모형식 : 공변량 cigar의 기울기는 처리에 대해 모두 동일

$$\hat{y}_{1j} = \hat{\beta}_0 + \hat{\alpha}_1 + \hat{\beta}_1 x_{1j} + \hat{\delta}_1 x_{1j} = 75.61 + 198.42 - 0.17\, cigar_{1j} - 2.48\,(cigar_{1j} \times trt)$$

$$= 274.03 - 2.65\, cigar_{1j}$$

$$\hat{y}_{2j} = \hat{\beta}_0 + \hat{\alpha}_1 + \hat{\beta}_1 x_{2j} + \hat{\delta}_2 x_{2j} = 284.52 - 3.01\, cigar_{2j}$$

$$\hat{y}_{3j} = \hat{\beta}_0 + \hat{\alpha}_1 + \hat{\beta}_1 x_{3j} + \hat{\delta}_3 x_{3j} = 75.61 - 0.17\, cigar_{3j}$$

● 치료법과 흡연량의 효과가 모두 음(-)이므로 흡연량이 증가하면 모든 치료법에서 산소운반능력이 떨어짐을 보여줌

- **cf : profiles of name of variables**

Sources	General Form of Aliasing Structure
intercept	Intercept + trt[3]
cigar	cigar + cigar*trt[3]
trt1	trt[1] - trt[3]
trt2	trt[2] - trt[3]
cigar*trt1	cigar*trt[1] - cigar*trt[3]
cigar*trt3	cigar*trt[2] - cigar*trt[3]

  - intercept는 trt3의 효과가 결합되어 있다는 의미
  - cigar는 cigar*trt3의 교호작용 효과가 결합되어 있음을 의미
  - trt1은 trt3과의 차이를 나타낸다는 의미
  - cigar*trt1은 trt1과 trt3에서의 담배소비량의 효과차이를 나타냄을 의미